STUDIES OF VOICE COMMUNICATION IN THE NAVY: SENTENCE PREDICTABILITY

by

Thomas G. Giolas
Harry S. Cooker
Joseph R. Duffy

University of Connecticut

SUBMARINE MEDICAL RESEARCH LABORATORY
NAVAL SUBMARINE MEDICAL CENTER REPORT NO. 653

Bureau of Medicine and Surgery, Navy Department
Research Work Unit M4305.08-3003D.02

Transmitted by:

J. Donald Harris

J. Donald Harris, Ph.D.
Head, Auditory Research Branch

Reviewed and Approved By:

Charles F. Gell

Charles F. Gell, M.D., D.Sc. (Med.)
Scientific Director
NavSubMedRschLab

Reviewed and Approved By:

Joseph D Bloom

J. D. Bloom, CDR, MC, USN
Officer-in-Charge
NavSubMedRschLab

Approved and Released By:

J. E. Stark

J. E. STARK, CAPT, MC, USN
COMMANDING OFFICER
Naval Submarine Medical Center

SUMMARY PAGE

THE PROBLEM

To investigate the predictability status of various sentence lists for use in evaluating hearing abilities of Navy personnel.

FINDINGS

The results indicate that sentence predictability plays an important role in the interpretation of the intelligibility scores obtained using sentence tests. Moreover, the predictability status of sentences was shown to vary so that low predictability sentences can be used to determine the message itself, while high predictability sentences can be used to determine one's ability to predict messages.

APPLICATION

The predictability status of sentences can be manipulated to provide additional evaluation of various hearing problems incurred by Navy personnel. Also, sentence lists with various predictability levels may be used as an additional tool for studying the ability to understand messages from fragmentary information such as in the case of message sending from hyperbaric or underwater environments which create distortions in the speech.

ADMINISTRATIVE INFORMATION

This investigation was conducted as a part of Bureau of Medicine and Surgery Research Work Unit M4305.08-3003D - Development of Auditory Screening and Acoustical Tolerance Standards for Submarine/Shipboard Personnel. This report has been designated as Submarine Medical Research Laboratory Report No. 653. It is Report No. 2 on this Work Unit and was approved for publication as of 23 February 1971.

The investigators of this project were furnished under ONR Contract with the University of Connecticut (No. N00014-67-A-0197-0001).

PUBLISHED BY THE NAVAL SUBMARINE MEDICAL CENTER.

ABSTRACT

The purpose of this study was to evaluate the predictability status of several sentence lists for use in hearing tests of Navy personnel. One hundred and eighty normal-hearing subjects, divided into six groups, listened to three sentence lists presented in one of six word-elimination conditions. The sentence lists were scored in terms of the number of words correctly predicted out of the number of words eliminated. The following results were obtained: (1) It was necessary to remove between 20 to 25 keywords for Sentence Lists B and D, respectively, before obtaining a significant score reduction. These lists were found to be highly predictable and quite similar with regard to their response to key-word elimination. (2) Considerably lower scores were obtained for Sentence List C than those obtained for Lists B and D. Furthermore, increased degradation through increased key-word elimination did not significantly affect the predictability scores. It was concluded that List C is considerably less predictable than Lists B and D. (3) The error curve for a Synthetic Sentence List was found to be extremely low and relatively flat, suggesting that this group of sentences has a negligible predictability value.

STUDIES OF VOICE COMMUNICATION IN THE NAVY: SENTENCE PREDICTABILITY

INTRODUCTION

Sentence lists have received renewed attention (Berger[1]; Giolas[2]; Harris, Haines and Meyers[3]; Jerger, Speaks and Trammell[4]) as measures of speech intelligibility. It has been argued that sentences present a more natural listening task than do the often employed monosyllabic word lists (Hirsh, Davis, Silverman, Eldert and Benson[5]) and take advantage of crucial parameters utilized in understanding connected speech. The Central Institute for the Deaf (CID) Sentence Lists (Silverman and Hirsh[6]) have been suggested as one possible message for clinical use. This group of sentence lists was developed to be representative of colloquial speech and a close relationship has been demonstrated between scores obtained using these sentence lists and a sample of continuous discourse (Giolas[2]). Furthermore, the lists are readily available in written form (Davis and Silverman[7]), require little administration time, are purported to include ten equivalent lists, and are easy to score. On the other hand, it has been shown that speech intelligibility scores obtained with these sentences are rather high under low levels of frequency distortion (Giolas[2]). One possible explanation of these results is that the sentences comprising this series are highly redundant. That is, a large portion of the sentence can be predicted on the basis of hearing only a small portion. It is believed that this group of sentences may not comprise a sufficiently sensitive test to detect subtle differences in speech intelligibility which may exist under minimally degraded conditions. Therefore, the purpose of the present study was to assess the predictability status of selected CID Sentence Lists, as well as selected lists from two other series (Harris, Haines, Kelsey and Clack[8], Jerger, et al[4]). Theoretically, a low predictability status would be desirable if sentence intelligibility scores are to be more dependent on speech discrimination ability than on the influence of contextual clues.

METHOD

Speech Samples. Scores obtained on CID Sentence Lists B and D, Revised CID Sentence List C and ten third-order synthetic sentences, constructed by Jerger, et al[4] were compared (see Appendix A for these four sentence lists). CID Sentence List B was selected in order to look at the possible predictability factor found with this list in a previous study (Giolas[9]). CID Sentence List D was employed so that some observations could be made concerning list equivalency in this series. Revised CID Sentence List C consists of an altered version of the original CID List C in which the sentences were modified to provide greater homogeneity of sentence length, while maintaining the colloquial speech criterion. This list was included in order to investigate the effect of sentence length on sentence-imbedded word predictability. The synthetic sentences were included to provide a

method for assessing the validity of the procedure employed to measure word predictability in sentences. A valid procedure would yield a low predictability score in that contextual clues in synthetic sentences are minimal.

Word Predictability Procedure. For purposes of this study, word predictability is defined as the property of a sentence which permits the prediction of a missing word(s). A modified version of the "cloze test* procedure" was employed to measure word predictability within a sentence structure. The procedure typically involves replacing a portion of a written message by a blank space. It was originally developed to measure the comprehensibility of a written message by requesting the subject to predict the missing words (Taylor[10]).

In the present study, an auditory cloze test procedure was employed to measure the predictability status of sentence lists typically used in speech intelligibility testing. It was assumed that a group of sentences containing highly predictable, sentence-imbedded words would yield high speech intelligibility scores as a function of message prediction rather than message reception.

All sentence groups were recorded on magnetic tape by a male speaker using an Ampex Model 300 Tape Recorder and a high quality microphone. Each sentence list was subjected to six conditions of word elimination. Selection of words to be eliminated was conducted on a random rather than a periodic basis to avoid systematic elimination of content or functional words. While there is some evidence that cloze test scores may vary with the elimination of different words (Greene[11]), little has been established as to the nature and extent of the score changes associated with different procedures for word deletion. A random procedure was employed, since it is the most effective method of avoiding any systematic bias. CID Sentence Lists B and D and Revised CID Sentence List C typically are scored in terms of a key-word count (Davis and Silverman[7]); consequently, only keywords were considered for elimination. Since the suggested scoring procedure for the Synthetic Sentences involves a total word count, all words were considered for elimination. Word elimination was accomplished by carefully cutting out the selected words and replacing them with an identical amount of leader tape. Six word-elimination conditions were established: Condition I, 10 words eliminated; Condition II, 15 words eliminated; Condition III, 20 words eliminated; Condition IV, 25 words eliminated; Condition V, 30 words eliminated; and Condition VI, 25 words eliminated. Condition VI was identical to Condition IV, except that a different group of words was eliminated. This condition was included in order to obtain some indication of the influence of word selection on the scores obtained. All treated speech samples were re-recorded to eliminate splice noise and to insure precise presentation.

TEST ADMINISTRATION: The tape recordings were played on an Ampex

*cloze test -- The term "cloze" was coined from the psychological concept of closure. Thus, an auditory "cloze" procedure implies that certain words are deleted from a spoken message and listeners are requested to close the gaps by filling in the missing words.

Model-300 Tape Recorder, through a high-fidelity amplifier to thirty matched TDH-39 earphones. The sentences were presented to the left ear at a comfortable loudness level in a relatively quiet setting. One hundred and eighty normal-hearing male listeners were divided into six groups of 30 subjects each. Each group listened to all four sentence lists under one of the word elimination conditions. Answer sheets and pencils were distributed to the subjects who were seated in tablet arm-chairs. They were requested to write the missing words in the appropriate blanks provided in the answer sheet. The answer sheet included the same amount of blanks for each sentence as words in the sentence. A series of x's were inserted in the blanks for the words not eliminated. The specific instructions were as follows:

> The purpose of this test is to see how well you can tell what the missing words are in a group of sentences. In other words, you will hear a group of sentences in which certain words have been eliminated. You are to guess what the missing words are. Each sentence will be preceded by its number. Listen to the complete sentence and then write in the missing word or words in the blanks provided for you on the answer sheet. The blanks containing x's indicate where in the sentence you actually heard a word and the empty blanks indicate the missing words you are to guess. You must fill in each blank with a word. Do not leave any blanks. PRIOR TO PRESENTATION OF SYNTHETIC SENTENCES: The last group of sentences will have very little meaning. Continue to guess at the missing words.

The subjects were given 60 seconds to complete the blanks for each sentence, but the full 60 seconds were seldom needed. The sentences were scored in terms of the number of words accurately predicted. Approximate spellings were accepted as correct. Unfilled blanks were counted as incorrect.

RESULTS AND DISCUSSION

Differences Between Conditions. Means and standard deviations of scores obtained for the four sentence lists studied within each word-elimination condition are summarized in Table I. The means recorded represent the number of words correct out of the total number of words eliminated. As indicated in the procedures, the number of words eliminated increased with each condition and, consequently, each mean must be viewed in relation to its own reference point. For example, the mean of 6.60 listed for Condition I represents 6.60 words correctly predicted out of ten words eliminated, while the mean for Condition II represents 7.63 words correctly predicted out of 15 words eliminated, etc. For purpose of comparison, the means have been converted into percentages and plotted in Figure 1. A zero word-elimination was not included, for it was assumed that this condition would yield a score of approximately 100%.

Z values for differences between proportions were computed between Condition I and all subsequent Conditions for each list. A significant Z was obtained between Conditions I-III, and I-V for CID Sentence List B (.05) and between Conditions I-III, and I-V for CID Sentence List D. This suggests that it is necessary to remove between 20-25 of the keywords before a significant score reduction takes place. It was concluded that Lists CID-B and D may be considered quite similar with respect to the effects of key-word elimination as accomplished in this study. On the other hand, Revised CID Sentence List C is affected quite differently

TABLE I - Summary of Means and Standard Deviation for Sentence List Scores Under Word-Elimination Conditions

Word Elimination Condition		Sentence Lists			
		CID List B	CID List D	Revised CID List C	Synthetic List
I (10 words)	Mean	6.60	6.03	2.53	.466
	SD	.129	1.423	.212	.118
II (15 words	Mean	7.63	7.36	5.66	.560
	SD	1.94	2.24	1.84	.185
III (20 words)	Mean	11.13	6.90	7.43	1.900
	SD	2.46	2.54	1.80	.310
IV (25 words)	Mean	6.43	8.70	3.93	1.70
	SD	2.26	2.24	1.79	1.25
V (30 words)	Mean	5.86	8.56	4.66	1.13
	SD	2.31	2.17	2.14	.047
VI (25 words)	Mean	6.26	7.96	3.03	2.76
	SD	1.85	2.62	1.30	.289

by key-word elimination. As can be seen in Figure 1, there is a considerably lower score for the minimal key-word elimination condition than for CID Sentence Lists B and D. Furthermore, increased degradation through increased key-word elimination did not significantly affect the predictability scores for subsequent conditions. No significant Z values were obtained between Condition I and all other Conditions for Revised CID Sentence List C. It was concluded that Revised CID Sentence List C is considerably less predictable than CID Sentence Lists B and D in that minimal degradation (10 key-word elimination) and increased degradation produced no significant change in predictability scores. It is suggested that the "filler" words used to equate the length of the sentences comprising the list were more difficult to predict and, consequently, reduced the list's predictability status.

The error curve plotted in Figure 1 for the Synthetic Sentence List (SSL) is relatively flat. While a significant Z (.01) was obtained between Conditions I-III and I-IV, the numerical difference is

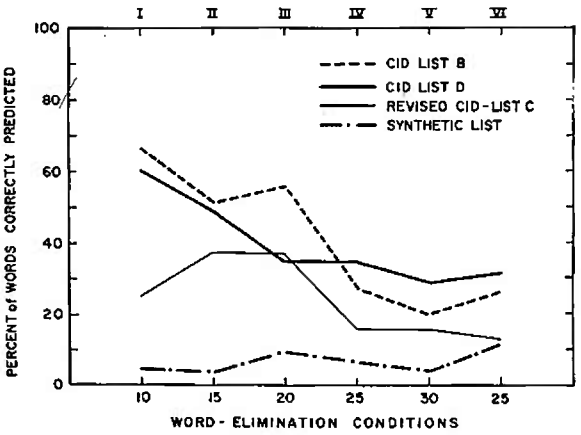

Fig. 1. *Comparison of Four Sentence lists (CID Lists–B and D, Revised CID List–C, and a third order Synthetic Sentence List) in terms of Percent of Key Words Correct for Six Word–Elimination Conditions.*

quite small and inconsequential. It was concluded that the SSL group has a negligible predictability value. A low, relatively flat curve for the SSL group was expected in that the anticipated predictability value for non-meaningful sentences is quite low. In a sense, the extremely low scores obtained for the SSL group for all conditions lend support to the use of word elimination procedures to study word predictability.

Theoretical and Obtained Scores: The predictability status of the four sentence lists investigated in this study is also demonstrated in Figure 2. Figure 2 graphically illustrates the relationship between theoretical scores and the scores actually obtained for each word-elimination condition. The theoretical curve represents points at which speech intelligibility scores would fall if the predictability status of each sentence list studied were zero; it is based on the assumption that all eliminated words would not be identified correctly. The points plotted on the obtained curve were derived by combining the theoretical score and the number of words correctly predicted in each condition. For example, 10 of the fifty keywords were eliminated in Condition I. If none of the eliminated words was predicted, the theoretical score would be 40. However, the actual mean number of eliminated words correctly predicted for Condition I was 6.60, yielding an obtained score of 46.60 [theoretical score (40) and number of words predicted (6.60)]. The curves plotted in Figure 2 include the theoretical and obtained scores for each condition. Confidence intervals computed (.01 levels) for the mean number of words correctly predicted at each condition were extremely small (see Table II). These findings indicate there is a real separation between the theoretical and obtained curves.

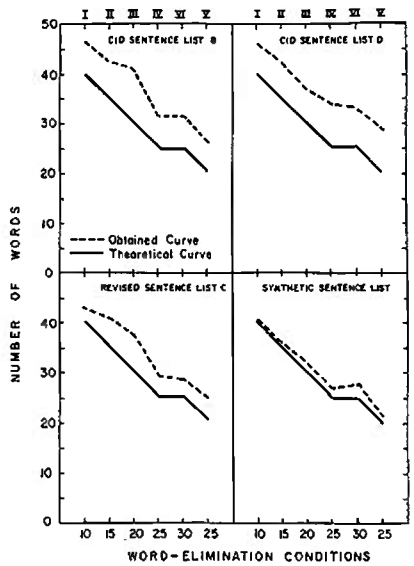

Fig. 2. *Comparison of Theoretical and Obtained Scores for Four Sentence Lists (CID Lists–B and D, also Revised CID List–C, and a third order Synthetic Sentence List) under Six Word–Elimination Conditions.*

TABLE II - Means and .01 Confidence Intervals For All Sentence Lists at Each Condition

Word Elimination Condition		Sentence Lists			
		CID List B	CID List D	R-CID List C	SSL
I (10 Words)	Mean	6.60	6.03	2.53	.466
	Confidence Interval	.633	.712	.105	.057
II (15 Words)	Mean	7.63	7.36	5.66	.560
	Confidence Interval	.973	1.12	.924	.090
III (20 Words)	Mean	11.13	6.90	7.43	1.900
	Confidence Interval	1.23	1.28	.904	.154
IV (25 Words)	Mean	6.43	8.70	3.93	1.70
	Confidence Interval	1.135	1.12	.899	.627
V (30 Words)	Mean	5.86	8.56	4.66	1.13
	Confidence Interval	1.16	1.09	1.075	.129
VI (25 Words)	Mean	6.26	7.96	3.03	2.76
	Confidence Interval	.930	1.31	.651	.137

Differences Between Lists: Mean number of words correctly predicted over all word elimination conditions along with the t-values computed between these means are listed in Table III. CID Sentence Lists B and D yielded extremely similar overall means which support the earlier conclusion that these two lists are quite similar with respect to sentence-imbedded word predictability. Sentence intelligibility scores typically obtained with either of these lists will be inflated by approximately eight key words or 16%, because of a listener's ability to predict words not heard. The mean number of words correctly predicted overall conditions for Revised CID Sentence List C was 4.846 or approximately 10%. This is a lower mean (.01) than the means obtained for CID Sentence Lists B and D, but still represents an appreciable inflation of an intelligibility score resulting from contextual cues. The overall mean of 1.153 obtained for the the SSL group further emphasizes that this group of sentences has the lowest predictability status of all lists studied. Furthermore, intelligibility scores obtained with these sentences

TABLE III - Differences in Mean Scores Between Test Lists Over all Conditions

Sentence Lists	Means	Differences	Standard Error of Differences	t
CID-B CID-D	7.533 7.513	.020	.274	.072
CID-B R-CID-C	7.533 4.846	2.687	.184	*14.744
CID-B SSL	7.533 1.153	6.380	.219	*29.132
CID-D R-CID-C	7.513 4.846	2.667	.234	*11.397
CID-D SSL	7.513 1.153	6.360	.192	*33.125
R-CID-C SSL	4.846 1.153	3.693	.187	*19.748

*Significant at the .01 level of confidence

would be least contaminated by sentence-imbedded word predictability.

Word Selection: In order to assess the influence of specific words eliminated on the predictability scores obtained, a comparison was made between Conditions IV and VI. As you will recall, Condition VI was identical to Condition IV, except that a different group of words was eliminated. As can be seen in Table III, the means for these two conditions for all sentence lists are extremely similar and Z values obtained for the differences between these means were non-significant (.01).

It was concluded that specific words selected for elimination did not have an appreciable effect on the predictability of the sentence lists studied.

CONCLUSIONS

The results of this study suggest that sentence predictability plays an important role in the intelligibility score ob-

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Speech Intelligibility						
Sentence Predictability						
Hearing Testing						

DD FORM 1 NOV 65 1473 (BACK)

(PAGE 2)

UNCLASSIFIED

Security Classification